DOCUMENT RESUME

ED 206 177 FL 012 481

AUTHOR Shohamy, Elana
TITLE The Stability of Oral Proficiency Assessment on the Oral Interview Testing Procedures.
PUB DATE 81
NOTE 37p.: Paper presented at the Colloquium on Validation of Oral Proficiency Tests (3rd, Ann Arbor, MI, 1981) and at the TESOL Convention (Detroit, MI, March 3-8, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Communication Skills; Correlation; Hebrew; *Language Proficiency; Language Styles; Language Usage; Language Variation; Second Language Instruction; *Second Language Learning; Semitic Languages; Social Environment; *Speech Skills; *Testing; Test Reliability; Test Validity
IDENTIFIERS Foreign Service Institute DC

ABSTRACT

While the goal of language teaching has been the development of practical communication, most language tests have stressed mastery of discrete linguistic skills. Oral language performance-based testing requires test-takers to use and perform in the language in actual communicative situations. The variations in communication created by the interaction between language and social setting have direct implications for testing oral communication. The extent to which communication variations manifest themselves in testing situations and the extent to which the assessment of oral proficiency is affected by these variations is unclear. A study was undertaken to determine the stability of the assessment of oral proficiency through four oral interview tests. Each oral interview was administered on different occasions by different interviewers, using different speech styles and topics. There was little effect on communication scores due to the variance of occasion or interviewer. However, the assessment of oral proficiency did change significantly from an "interview" style to a "reporting" style, when subjects spoke on different topics. Although there is a lack of stability and low reliability and validity present in the FSI Oral Interview, the test does provide an excellent setting for a variety of tests for different communicative variations. (Author/JK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED206177

THE STABILITY OF ORAL PROFICIENCY ASSESSMENT ON THE ORAL INTERVIEW TESTING PROCEDURES

by

Elana Shohamy
Stanford University/U.C., Berkeley

A preliminary version of this paper was presented at the Colloquium on Validation of Oral Proficiency Tests (3), Ann Arbor, Michigan, 1981, and at the TESOL Convention, Detroit, Michigan, 1981.

FL 012 481

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Elana Shohamy

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

The paper first describes the complexity of measuring oral proficiency in communicative situations, due to the large number of linguistic and social variables. It then reports on a study which examined the stability of the assessment of oral proficiency through four oral interview tests, when each oral interview was administered on different occasions, by different interviewers, using different speech styles and topics. Results from the analysis of variance indicated that a different speech style and topic resulted in significant differences in the means of the tests. Results of the correlational analysis between scores of the tests pointed to low reliability especially when two variables (i.e., occasion and tester), interacted.

PERFORMANCE-BASED ORAL TESTS

It has been noted that language testing should follow closely trends in language teaching. However, for a long time there seems to have been a discrepancy of large proportion between the two domains (Spolsky, 1978; Jones, 1977); while the goal of language teaching was the development of practical communication, most language tests stressed mastery of discrete linguistic skills. Recently, though, there have been many calls and specific suggestions for testing communication and other pragmatic skills (Morrow, 1977; Can le and Swain, 1980; Carroll, 1980). Thus, oral testing, a domain in language testing which was badly neglected in recent years, is receiving major attention. The most advocated approach to testing oral communication is to require the test-takers to use and perform in the language in actual communicative situations. Canale and Swain (1980), contend that communicative testing must be devoted to not only what the learner knows about the second language and about how to use it, but also to what extent the learner is able to actually demonstrate this knowledge in a meaningful communicative situation. Jones (1977), claims that if testing

is to provide useful information about functional language ability, it seems fitting that language tests should approximate real tasks of communication, and he calls for performance tests as the better choice for assessing real communication skills.

Clark (1975), calls for the use of direct performance tests for assessing oral proficiency. Such tests require the test taker to use the language in actual 'real life' situations. He, thus, advocates the use of the oral interview for such purposes, because it closely resembles real life situations, and thus, has high face validity. He points out, though, that less direct tests, termed by him "semi-direct", can be acceptable for testing speaking proficiency as long as they prove to correlate very highly with direct tests, indicating that similar information can be obtained from the more efficient indirect tests as from the direct tests. Morrow (1977), focuses on analysis and procedures for testing communication. He suggests both performance and non-performance tests for testing communication, but cautions that non-performance tests measure only a limited aspect of

language competence. He calls for tests that offer the speaker the opportunity for spontaneous production and for actual performance for operating the language system in authentic, communicative settings. He contends that such tests would be based on activities which the candidate should recognize as relevant to the use he is likely to make of the target language, and lists tests such as answering questions, asking questions, interaction participation and narrations to assess oral communicative ability.

The call by many researchers for performance-based oral communicative assessments is relatively new in the testing field and clear distinction is being made here between tests of "knowledge" and tests of "performance". While in "knowledge" tests, proficiency in language is determined on the basis of knowing about the language without the need to prove it in communicative situations, "performance" tests require the test-taker to apply the knowledge by actually using the language in such situations.

While "knowledge" tests specify in exact terms the elements one is expected to know, "performance"

tests define knowledge in terms of expected behavior in communicative situations. While the "knowledge" tests focus on correctness of grammar and vocabulary, "performance" tests take into account the application of these elements within a communication activity, and thus consider also such criteria as saying the right thing at the right time, to the right person. While the "knowledge" tests are believed to be reliable, "performance" based tests are frequently referred to as subjective, and unreliable, and thus are believed by many to provide inaccurate assessment of oral proficiency.

As a result, there was a need for research on the reliability and validity of such tests. Research on rater reliability, which is the degree of agreement among independent raters regarding the assessment of test-takers, was found almost consistently to be high among trained raters (Hinofotis, 1976; Clifford, 1978; Shohamy, 1978), as well as among naive, untrained raters (Shohamy, 1980). Validity of performance-based tests is being examined through construct validity (Bachman and Palmer, 1981; Clifford, 1978; Brütch,

1979; Henning, 1981), and through concurrent validity, i.e., correlating scores obtained from oral communicative tests with scores obtained on other skill tests (Shohamy, 1978; Hinofotis, 1976), and with other less direct oral tests, such as the MALT (Clifford, 1977; Brütch, 1979). It is worth noting that most of this research uses the FSI Oral Interview, or version of that test, as the oral communicative test.

MEASURING COMMUNICATIVE PERFORMANCE

In spite of the research available on reliability and validity of communicative oral tests, there are still many unresolved measurement problems. These are due to the difficulty and complexity of defining the communicative construct, since it is not entirely clear what constitutes oral proficiency in a communicative situation and what the specific variables are which interact in that construct.

Without knowing the specific variables which are involved in the oral communication act and their relative contribution to communication makes it difficult to standardize communicative tests, since we do not know how a change in the communicative

situation may affect oral proficiency and consequently its assessment.

Sociologists have attempted to define communication and to isolate its variable. They introduced theories indicating that communication includes not only grammaticality, but also contextual sociologistic elements (Hymes, 1971, 1972). They claimed that considerations for who is saying what, to whom, in what circumstances and under what conditions are important considerations in communication. Variables such as the topic, context, channel, the status of the encoder and decoder, age, etc., have a bearing on the choice of codes, styles and registers to be used. Hymes (1967) introduced the model of interaction of language and social setting and specified the components of communication to be: the setting -- time and place; scene -- psychological or cultural setting; participants -- speaker and hearer and their social roles to each other; topic -- what the message is about; key -- mock, serious; channel -- oral, written; norms of interaction and interpretations -- loudness of voice, when to interrupt, and how violations

of norms of interaction are interpreted; and fendre -- causal speech, prayer, form letter.

Morrow (1977), suggests an approach to oral communication which isolates nine variables which have a bearing on any speech act. The variables he suggests are: setting, topic, function, modality, pre-supposition, role, formality, status, and mood -- all believed to interact in the communication process.

Another view of the communicative act is provided by Joos (1967) who suggests a classification of speech styles using the criterion of formality; this criterion tends to subsume subject matter, the audience, the mode of course, and the occasion. Joos recognized five different levels of formality, each implying different forms of speech to fit separate functions: (1) oratorical, or "frozen", (2) deliberative, or formal, (3) consultive, (4) causal, and (5) intimate.

As can be observed from these classifications, communication is a very involved construct containing language and non-language variables and it is clear that combinations among the variables create variations in communication. However, it is not clear which

components are important in each communication act. Walters(1978) indicated that contextual factors such as setting, topic, and sex, age and race of participants are the most salient in their effect on variations, but there is much less research on variation in the expression of other communicative functions. Canale and Swain (1980), state that it is reasonable to conclude that relatively little is known about how social context and grammatical form interact.

Hudson (1980), states that if speech is skilled work, the same is true of other aspects of social interaction in face to face communication. Thus, some people are better at social interaction than others; success in speech varies considerably according to its functions and other aspects of the situation. However, it is not obvious how successful speech should be measured since it is required to include general skills for social interaction plus all the specifically linguistic skills concerned with the use of linguistic items.

It is clear from this review that the interactions between language and social setting create

variations in communication. These variations have direct implications for testing oral communication, for it is in testing that we have to account for the variations to assure that we can replicate the results with a high degree of accuracy. However, it is not clear to what extent these communication variations manifest themselves in testing situations and to what extent the assessment of oral proficiency is affected by these variations. One procedure that can be used to examine it is by administering several communicative oral tests where each administration is different than the other in one communicative variable and then compare the assessment of oral proficiency obtained from the various administrations. By changing the variables in two identical oral communicative situations, it should be possible to observe whether the interaction of language and social setting manifests itself significantly in the assessment of oral proficiency. Such a procedure may also help in estimating the extent to which social variables affect oral proficiency. If significant differences are found in the assessment of oral communication tests, which are based on such variations,

there will be evidence that these variations manifest themselves in the particular test used and that the test does not account for these differences. If, on the other hand, no significant differences are obtained in the assessment of the scores of such communicative tests, there will be evidence that these variables do not affect the assessment of oral proficiency on the particular test, and that the test used may adequately assess these variations.

THE PURPOSE OF THIS STUDY

The study reported here describes such a procedure. It examines the differences in the assessment of oral proficiency when the testing occasion, the tester, the speech style, and the topic were each introduced independently into an oral communicative test. If the introduction of these variables significantly affects the assessment of oral proficiency, there will be evidence for lack of stability of the oral communicative test, in this case, the Oral Interview.

It is claimed in this paper that since communicative testing situations involve a large set of variables, it is necessary to examine their relationship and effect on the assessment of oral

proficiency. To what extent is the assessment of oral proficiency stable along these variables? Results from such an examination have a major impact on the accuracy of the assessment of oral proficiency.

It is also claimed that since there are many different communicative situations which a second language speaker is likely to encounter, it is necessary to examine whether the oral proficiency of a speaker changes significantly in each of these communitative situations, and if certain oral communicative tests are valid predictions of assessment of other communicative speaking situations.

The Oral Interview (OI) was the communicative test chosen for this research, since it is a performance-based oral communicative test which is widely used and researched; it offers an approach to language testing, whereby the test-taker actually communicates with a trained interviewer in the target language and as such, it is a performance-based communicative test.

Wilds (1975), states that a serious issue with the FSI OI is the stability of performance with different sets of interviews. She claims that

there are two legitimate and interesting questions that FSI cannot answer.

(1) Does the proficiency of the speaker of a foreign language fluctuate measurably from day-to-day?

(2) Does his performance vary with the competency and efficiency of the examiner?

Another problem with the OI is that it represents a limited domain of communicative speech styles -- an interview, which is not the most frequent speech style a foreigner is likely to encounter in the target culture. It is of interest then to find out if assessment through the OI reflects an assessment based on other speech styles.

Thus, the two main questions that his research attempts to answer are:

(1) How stable are the scores which are obtained on an Oral Interview, when communicative variables such as the occasion, and the interviewer change? This is essentially a reliability

question since it deals with consistency of scores within the same test.

(2) How accurately does the assessment of oral proficiency obtained from a communicative test such as the Oral Interview, which is based on interviewing reflect assessment of oral proficiency based on a 'reporting' task? This is a validity question, since it deals with variations between the OI and a slightly different version of the test.

PROCEDURE

Forty students of Hebrew at the University of California, Berkeley, underwent four communicative oral tests. Students ranged in proficiency from beginning to advanced.

Test 1 was an Oral Interview, conducted by the researcher, a teacher in that program. Its duration was ten to twenty minutes and followed Lowe's (1976) four steps -- warm-up, level check, probes, and wind-up. Test 2 was the identical oral interview, but

was conducted four weeks later. Test 3 was the same oral interview test as 1 and 2, except that it was conducted by a different interviewer, an advanced male Hebrew student whose first language was English. His Hebrew is highly proficient, and he was trained by the researcher to conduct oral interviews. In test 4, the test-takers were supposed to report orally in Hebrew to the interviewer about an article about the risks of smoking that they had just read. The test lasted approximately five minutes. This activity was selected since it was believed that reporting is a task which is performed frequently in real life.

Thus, test 4 differed from the first three tests in speech style, and topic. All other variables -- the length of test, the setting,etc., were controlled as much as possible. Tests 2, 3, and 4 were administered in a mixed order to control sequency effects, and were administered within a period of ten days.

Insert Figure 1 about here

All four tests were taped and later rated independently by two native Hebrew raters using the oral proficiency Hebrew scale (Shohamy, 1978), which defined oral proficiency in terms of grammar, vocabulary, fluency and pronunciation. Overall proficiency was rated using the FSI rating scale.

The maximum rating was 5, the lowest 0. Whenever raters were in doubt, they were encouraged to use pluses and minuses to increase the variance. Each tape was assigned a code number so that the identity of the speaker would not be disclosed to the raters, in order that they would not view the tests as four tests by the same person, but rather as independent tests.

RESULTS

Inter-rater reliability, which was a pre-requisite for any of the subsequent analyses, was computed as .93. It was high enough to allow for the use of the average score of the two raters in all the computations.

Three-way analysis of variance was computed on the scores of the four oral tests to examine the

differences in the means.

Insert Table 1 about here

Table 1 displays the means and standard deviations of the four tests, and Table 2 presents the analysis of variance table. It indicates a significant main effect at the .01 level for overall oral proficiency as well as for grammar, vocabulary, and fluency, but not for pronunciation. It is clear from Table 1 that while tests 1, 2, and 3 were closer to one another on the means, test 4 had a considerably lower mean.

Insert Table 2 about here

Analysis of the contrasts between pairs of means (Table 3) using the Tukey contrast formula, revealed significant differences between tests 1 and 4, 3 and 4 on overall oral proficiency; between tests 1 and 4, 3 and 4 on grammar; between tests 1 and 4, 3 and 4 on vocabulary; and between tests 1 and 4 on fluency.

The other contrasts were not significant at the .01 level, but nevertheless, indicate a trend towards significance. No significant differences were obtained on any of the means for pronunciation. In fact, the means on pronunciation seem relatively stable on the four tests (3.15, 3.15, 3.37, 2.95). The significant difference obtained with test 4 suggests that the different speech style and topic used on that test affected the assessment. Although the differences between the other tests were not significant, it is well worth noting that the mean scores of test 3 were consistently higher than the means of test 2 on the various sub-components of oral proficiency and that the mean scores of test 1 were consistently higher than the scores of test 2.

Insert Table 3 about here

Pearson Product-Movement Correlations were computed for pairs of tests to examine the relationship between students' scores on the various tests as well as the mutual variance among these tests. Table 4

displays these correlations and the amount of common variance (r^2).

Table 4 displays the correlations between the scores of students on the various tests. It shows that the highest correlation, r = .80, was obtained between the two tests which varied by occasion (tests 1 and 2), and the lowest correlation, r = .36, was obtained between the two tests which varied by occasion and by interviewer, (tests 1 and 3). The other correlations where r = .66 for tests 2 and 3, r = .71 for tests 2 and 4, and r = .44 for tests 3 and 4. The correlation between tests 1 and 4 are not reported since they were based on a very small number of students.

Insert Table 4 about here

It should be pointed out here that the two types of statistical analyses -- analysis of variance and the correlational analysis -- provide two types of information for different kinds of decisions. The analysis of variance gives information about the effect

of the variances on the assessment of communicative oral proficiency, while the correlation analysis provides information on the relationship between the scores of the various tests (r) and, thus, the probability (r^2) of an individual to obtain the same score when she/he undertakes the other test.

CONCLUSIONS

1. Fluctuations in the Oral Interview ratings due to occasion and interviewer (tests 1, 2, 3)

The lack of significant differences among the means of the first three tests which varied by the occasion and by the interviewer, provided evidence for little effect of these variables on communication scores. The introduction of a different testing occasion and a different interviewer into the Oral Interview communicative test did not affect the scores of oral proficiency and did not point to significant fluctuations of the scores of the three tests.

In spite of lack of significant statistical differences of the three tests, the trend and slight fluctuations obtained between some of the means have significant administrative implications. For example,

a score of 3.6 on test 1 may be rounded off to 4.0, while a score of 3.03 on test 3 most likely be rounded off to 3.0. Obviously such differences may result in different administrative decisions for the test-taker.

The trend towards significance of some of the pairs of means (especially tests 2 and 3 which varied on the interviewer), suggest that possibly by introducing other variables or combinations of variables, fluctuations in the assessment are likely to come out.

The correlation analysis is more relevant to decisions which are made about individuals based on such test scores. While the scores between tests 1 and 2 were relatively high (.81), the correlations between tests 2 and 3 dropped to $r = .66$, and dropped substantially lower when the two variables, occasion and interviewer (tests 1 and 3) interacted ($r = .36$). The squared correlation is considered the mutual variance of the tests and thus is an indication of the probability of an individual obtaining a similar score if she/he undergoes the other test. Thus, the squared correlations (r^2)

displayed in Table 4, as low as $r^2 = .13$, show low mutual variance for tests 1 and 3.

The results from the two analyses call for careful control and clear specifications of the oral interview procedure, in terms of interviewer administration and content of the Oral Interview, to minimize the error and maximize reliability of that test. There is also a need to further test the reliability of the Oral Interview with other social variables as well.

2. Fluctuations in oral proficiency assessment due to speech style and topic (test 4)

The significant difference between the means of test 4 and the means of the other three oral interviews indicate that the assessment of oral proficiency in communicative tests tends to change significantly from an 'interview' style to a 'reporting' speech style, on different topics. These differences suggest that speech style and topic tend to affect the assessment of oral proficiency. It is not clear from this study if the significant difference was obtained as a result of changing the speech style, the topic

or the interaction of the two.

The moderate common variance between test 4 and test 2 (r^2=.50), and the low common variance between test 4 and test 3 (r^2=.19), indicate that the Oral Interview may account somewhat for differences in speech style with the same interviewer but accounts very little for any variance when the speech style, topic, and interviewer change.

The findings suggest the need to further examine the validity of the Oral Interview with other speech styles and other topics; different speech styles may imply different language proficiencies. One of the major criticisms about the Oral Interview is that it employs an 'interview' speech event which is not a very frequent event in real life; a conversation, for example, where people interact on an equal basis is probably a more frequent event in real life.

IMPLICATIONS

The results obtained in this study are based on a relatively small sample. However, they provide some indication of lack of stability and low reliability and validity of the Oral Interview.

It is necessary to examine to what extent the Oral Interview accounts statistically for other variations in communication as well, and if there are significant differences in the assessment when other social variations are introduced. If the Oral Interview is not found to account for other communication variations, as in this study, then there is a need to design special tests to focus on these variations, or possibly a battery of tests for various sociolinguistics variations. Such tests should be designed to reflect as much as possible the 'real life' tasks and objectives that a person is most likely to encounter in communicative situations, tailored to his goals and activities. In the same way that we talk about teaching for special purposes, we may need to design language tests for special purposes. Differing student needs and different situations may require different types of oral proficiency tests.

Nevertheless, the Oral Interview test provides an excellent setting for a variety of tests for different communicative variations. The setting involving a communicator and a communicatee who interact in the target language has great potential for

testing a variety of communicative situations, and should not be limited to interviewing. Other communicative activities, registers, speech styles, etc., can be exercised through that test as well. At the same time, there is a need for drawing stringent guidelines and instructions about the procedure and content of the test to assure uniformity and maximum consistency of the assessment. Such guidelines should focus on a variety of linguistic and social variables, such as the setting in which an interview should be conducted, the register to be used, the atmosphere of the interview, type of questions, the topics, norm of interaction, personality traits of the interviewer, and more, but until we know the exact effect of the essential variables which interact in the communicative process, we need to be cautious about generalizations made from oral communicative tests which are based on a single testing procedure, on a single occasion, and on a single communicative situation. There is a need to rely on sociolinguistic theories and research to further examine reliability and validity of oral communicative tests to achieve generalizability of

these tests and thus, to obtain accurate estimation of the errors of measurement.

BIBLIOGRAPHY

Bachman, L., and Palmer, A. 1981. The construct validation of tests of communicative competence. Paper presented at the Colloquium on the Validation of Oral Proficiency Tests (3). Ann Arbor, Michigan.

Brütch, S. 1979. Convergent/discriminant validation of prospective teachers, proficiency in oral and written production of French. Doctoral presentation, University of Minnesota.

Canale, Michael and Swain, Merrill. 1980. Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics, I, 1, 1-47.

Carroll, B. 1980. Testing communicative competence. Pergamon Institute of English.

Clark, John L.D. 1975. Theoretical and technical considerations in oral proficiency testing. In: R.L. Jones and B. Spolsky (Eds.) Testing language proficiency. Arlington, Virginia, Center for Applied Linguistics, 10-28.

Clifford, R. 1978. Reliability and validity of language aspects contributing to oral proficiency of

prospective teachers of German. In: John Clark (Ed.) Direct testing of speaking proficiency. Educational Testing Service, Princeton, N.J., 191-210.

Henning, G. 1981. Multi-trait and multi-method validation, fit to a one parameter logistic model and multiple regression procedures in the validation of 15 subscale of oral language production. Paper presented at the Colloquium on the Validation of Oral Proficiency Tests (3). Ann Arbor, Michigan.

Hinofotis, F. 1976. An investigation of the concurrent validity of Cloze testing as a measure of overall proficiency in English as a second language. Doctoral disseration, Southern Illinois University.

Hymes, D. 1967. Models of the interaction of language and social setting. The Journal of Social Issues, 23, 2, 8-28.

Hymes, D. 1971. On linguistic theory, communicative competence and the education of disadvantaged children. In: M. Wax, S. Diamond, and F.O. Gearing (Eds.) Anthropological perspectives on education. New York, Basic Books, 51-66.

Hymes, D. 1972. On communicative competence. In: J.B. Pride and J. Holmes (Eds.) Sociolinguistics. Harmondsworth, England, Penguin Books, 269-293.

Hudson, R.A. 1980. Sociolinguistics. Cambridge University Press.

Joos, M. 1967. The five clocks. New York, Harcourt, Brace and World.

Jones, R. 1977. Testing: A vital connection. In: June Phillips (Ed.) The language connection: From the classroom to the world. The ACTFL Review of Foreign Language Education Series, Volume 9. Skokie, Illinois, National Textbook Company, 237-265.

Lowe, P. Jr. 1976. Handbook on question type and their use in LLC oral proficiency tests. Washington, D.C., CIA Language Learning Center. (Preliminary Version).

Morrow, Keith. 1977. Techniques of evaluation for noting syllabus. Centre for Applied Language Studies, University of Reading.

Shohamy, Elana. 1978. Investigation of the concurrent validity of the oral interview with Cloze procedure for measuring proficiency in Hebrew as a second language. Doctoral dissertation, University of Minnesota.

Shohamy, Elana. 1980. Further examination of validity of the oral interview. Paper presented at the Colloquium on Validation of Oral Proficiency Tests (2). TESOL, San Francisco.

Spolsky, B. 1978. Introduction: Linguistics and language tests. In: B. Spolsky (Ed.) Advances in language testing series: 2. Approaches to language testing. Center for Applied Linguistics, Arlington, Virginia, v-x.

Wilds, Claudia. 1975. The oral interview test. In: Randalle Jones and Bernard Spolsky, (Eds.) Testing language proficiency. Arlington, Virginia, Center for Applied Linguistics, 29-44.

Walters, J. 1978. Social factors in the acquisition of a second language. Paper read at the AILA Congress, Montreal. (Reference obtained from Canale and Swain, 1980, Approaches to language teaching).

FIGURE 1. Description of the four tests used in the study

Test	Interaction	Partici-pants	Occasion	Speech Style	Content
Test 1:	one to one	teacher student	Dec.	Interview	X
Test 2:	one to one	teacher student	Jan.	Interview	X
Test 3:	one to one	student student	Jan.	Interview	X
Test 4:	one to one	teacher student	Jan	Reporting	Y

TABLE 1. Mean and Standard Deviation of the Scores of the Four Oral Tests

	Test 1 n=15		Test 2 n=30		Test 3 n=28		Test 4 n=12	
	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
OVERALL	3.60	.63	3.03	.85	3.39	.73	2.66	1.12
GRAMMAR	3.39	.65	2.84	.86	3.18	.76	2.57	1.10
VOCABULARY	3.45	.64	2.85	.89	3.22	.66	2.48	1.00
PRONUNCIATION	3.15	.45	3.15	.57	3.37	.52	2.95	.82
FLUENCY	3.59	.65	3.07	.83	3.46	.71	2.87	.99

TABLE 2. Summary of Analysis of Variance in Overall Oral Proficiency

Source	SS	df	MS	F
between	77951.09	3	25983.69	3.79**
within	554813.61	81	6849.551	

** $\alpha \leq .01$

TABLE 3. Contrasts which were found to be significant between pairs of means, using the Tukey formula

	Overall	Grammar	Vocabulary	Pronunciation	Fluency
Tests	1,4	1,4	1,4	--	1,4
	3,4	3,4	3,4	--	---

TABLE 4. Pearson Product-Moment correlations (r) and (r^2) between the pairs of tests on overall proficiency

Tests	n	r	r^2	Source
1, 2	12	.80	.64	test re-test
1, 3	14	.36	.13	test retest with interviewer
2, 3	26	.66	.44	inter interviewer
2, 4	12	.71	.50	inter speech style and topic
3, 4	8	.44	.19	inter interviewer inter speech style and topic